THE SEMANTIC REEF: AN ECO-INFORMATICS APPROACH FOR MODELLING CORAL BLEACHING WITHIN THE GREAT BARRIER REEF.

Trina Myers[1], Ian Atkinson[2], Jeff Maynard[3]

School of Maths, Physics and Information Technology, James Cook University.
Trina.Myers@jcu.edu.au[1], Ian.Atkinson@jcu.edu.au[2]

Great Barrier Reef Marine Park Authority. J.Maynard@gbrmpa.gov.au[3]

ABSTRACT

The Semantic Reef project is an application of eco-informatics and represents the first time that semantic tools have been applied to the study of coral reef ecosystems. It aims to develop a platform that combines emerging Semantic Web, Artificial Intelligence (A.I.) and Grid Computing technologies to make unconnected datasets computer-understandable. The key goal is to improve our capacity to generate timely warnings of environmental conditions that are conducive to coral bleaching. The Semantic Reef model is a tool for hypothesis-driven research and problem-solving methods, which will allow for the maintenance and analysis of disparate data streams. The initial validation exercise, to test the accuracy of the model using sea surface temperature and community composition datasets (both static and dynamic data), is presented. It is the unequivocal scientific consensus that climate change is a great threat to the Great Barrier Reef (GBR). The efforts described here represent an eco-informatics application of new information technologies to the forecasting and monitoring of climate-change-related impacts on the GBR. It is likely that combining the advanced technologies described here with powerful ecological modelling will result in entirely new lines of enquiry, as well as an improved understanding and management of environmental systems. While coral reefs are the focus environment herein, the concepts will be applicable to many other disciplines as the field matures.

KEYWORDS:

Coral bleaching, Eco-informatics, Semantic Web, ontology.

INTRODUCTION

The use of environmental sensor networks to gather data in real-time across widely distributed areas is an expanding field, detailed by the Integrated Marine Observing System (2007). Applications of new technologies and processing systems are being trialled on the Great Barrier Reef (GBR) (Kininmonth et al., 2004). However, using and making real-time data accessible in a timely and useful manner to managers remains a critical issue for applications scientists. Myers et al. (2007) formulated an ecosystem ontology, which included a description of a vision for the Semantic Reef Project described here (Figure 1). The goal of the semantic reef is to automate the analysis and interpretation of real-time data across the GBR. Myers et al. (2007) also described a methodology on ontology engineering using the Web Ontology Language (OWL), and

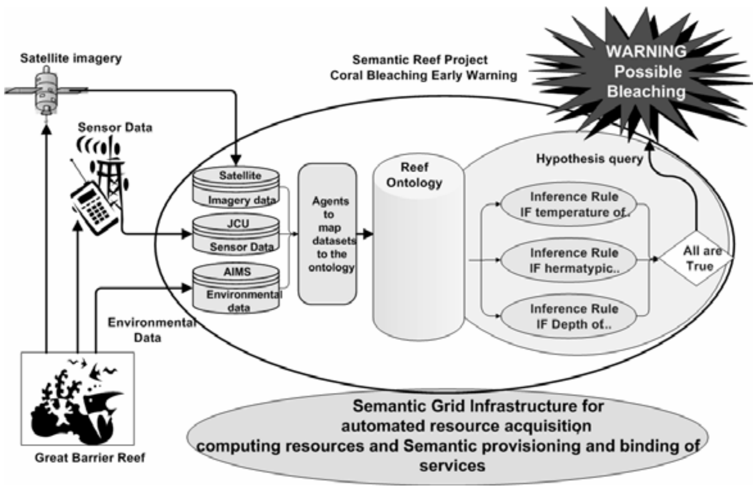

Figure 1 – The Semantic Reef architectural vision.

the use of Description Logics (DL) and inference rules to allow the computer to reach a conclusion based on the information specified. The projects' architecture is being developed to assist in hypothesis-driven research, problem-solving environments and ultimately in sensitivity analyses for the diverse number of disciplines that explore the correlations of interesting data to create new knowledge.

Successful validation of the workflow has been accomplished by adopting a reverse-hypothesis approach where the outcomes of the set inference rules within the reef ontology were evaluated against *in-situ* field observations from previous events. The full validation trials will include known causal factors of spatially extensive and severe coral bleaching such as temperature, light and water depth. Among other metrics, the Sea Surface Temperature (SST) and Degree Heating Days (DHDs) have been shown during the 1998 and 2002 bleaching events to be well correlated with the severity of bleaching responses (Berkelmans et al., 2004). In preliminary analyses, logical inference rules and Description Logics (DL) that mimic the aforementioned metrics of SST and DHDs were executed and results found to relate closely to those of previous research on the tolerance of corals to temperature changes.

There are a number of proposed initiatives to better monitor bleaching-related impacts, as well as develop management strategies, that work to enhance the natural resilience of reef systems (Marshall and Schuttenberg, 2006). Many proposed monitoring and management measures are only effective given early warning systems that can assess where bleaching and bleaching-induced mortality are likely to be most severe. The types of products developed by the Great Barrier Reef Marine Park Authority (GBRMPA), CSIRO, and the Australian Bureau of Meteorology under the ReefTemp initiative make spatial predictions about bleaching severity but do not currently take into account the great number of variables that can contribute to the bleaching response (Maynard et al., 2007). The models and tools described here will expand upon the current initiative by providing a method to automate the processing of a number of data-streams as well as determine the relative importance of different physical and ecological variables in eliciting the bleaching response. In this way, the resulting tools can improve the ability of managers to monitor and mitigate a major impact to the reef but also help to elucidate knowledge gaps that can inspire highly relevant research in the future.

Following a brief description of current methodologies used in the research on the coral bleaching phenomenon, the background and architecture of the Semantic Reef project is presented. Specifically, there is a focus on examining various methods and indices used on the 1998 and 2002 mass bleaching events and replicating them within an ontology. Secondly, we will briefly illustrate the validation process which involves imitating these methods and using description logics and inference rules to determine the accuracy of the ontology. Finally, the successes found in this stage of the project's implementation and a discussion of future possible applications using the problem solving workflow shown in Figure 1.

BACKGROUND - CURRENT RESEARCH METHODOLOGIES AND MATERIALS FOR CORAL BLEACHING

Two major coral bleaching events occurred in the GBR and Coral Sea in the late summer (February and March) of 1998 and 2002. Bleaching severity during each event was assessed using an underwater video survey technique, popularized by English et al. (1997), at 14 sites on the central GBR. These sites were used to explore the relationship between accumulated thermal stress and bleaching severity. Four of these sites (Faraday Reef, Kelso Reef, John Brewer Reef and Magnetic Island Reef) were re-surveyed in 2002 to evaluate changes in the relationship between thermal stress and bleaching severity between events. Historical records of the temperature regime at each site were obtained using remotely sensed sea surface temperatures (SST) from the satellite platforms of the U.S.

National Oceanic and Atmospheric Administration (NOAA) (Gleeson and Strong, 1995). The selected reefs showed significant levels of bleaching in both years, even though the strength and spatial distribution of thermal anomalies in the GBR lagoon differed between 1998 and 2002 (Berkelmans et al., 2004).

Four thermal stress indices were used to describe the thermal anomalies experienced by each study site during the 2002 summer; magnitude of SST anomaly (SST+), maximum temperature over any 3-day period ($3d_{Max}SST$), degree heating days (DHD) and heating rate (HR) (Maynard et al., 2007). The SST anomaly is calculated for each grid cell as the number of Celcius degrees above the long-term mean temperature (LMST) observed for that month. The temperature anomalies visualised range from +0.1°C to +5°C. The $3d_{Max}SST$ (Berkelmans, 2002) represents the peak temperature experienced during any 3-day period at each site. The SST+ is calculated as,

$$SST+ = 3d_{Max}SST - LMST$$

where the LMST is the long-term mean summer temperature at the site for the ten year period prior to the event. This index is indicative of the strength of the anomaly and is relative to the historical summer temperature regimes at each site.

The DHD and HR are indices that describe the accumulation of thermal stress (Maynard et al., 2007). Specifically, the DHD value is the summed positive deviations of daily average sea surface temperatures ($T_{Heating}$) from historical summer mean temperatures ($LMST$).

$$DHD = \sum(T_{Heating} - LMST)$$

BACKGROUND – THE SEMANTIC REEF PROJECT

The workflow (Figure 2) of the Semantic Reef shows that a range of existing technologies are being employed including Compute agents and Semantic Web technologies to allow for virtually-organised operation and the management of knowledge bases. This Semantic Grid

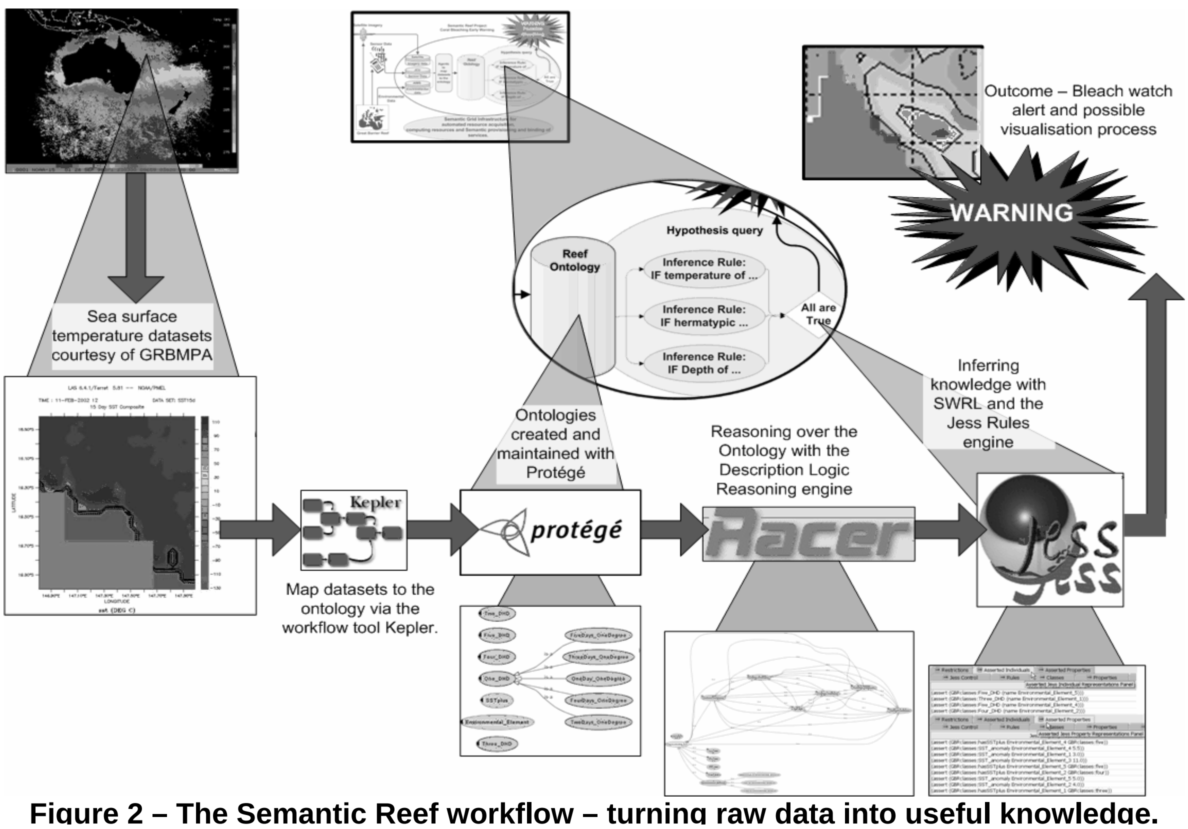

Figure 2 – The Semantic Reef workflow – turning raw data into useful knowledge.

platform can access many datasets simultaneously and process them using the agents-based technologies. The workflow begins with an agent that processes the lowest level data from several sources (e.g. sensor data, satellite SST data, etc). The agent will then map the analysed data to an ontology for hypothesis testing. Ontologies are the foundation of Semantic Web technologies and can be defined as descriptions to represent abstract or specific concepts, that is, intentions, beliefs, objects and feelings. These descriptions contain explicit specifications, terms and relationships with formal definitions, axioms and restrictions that constrain the interpretation to create, share and re-use computer understandable knowledge (Guarino, 1997). These Semantic technologies make raw data and information “understandable” to a computer, thereby enabling it to make intelligent decisions based on inference rules and description logics (Antoniou and van Harmelen, 2004).

Agent-based technologies are adaptive software programs that can perform proactively and/or reactively in dynamic environments (Jennings, 2001). In this application, we use the agent-oriented workflow software Kepler (Atkinson et al., 2007) to automatically process raw data, and pass the results onto the ontology. An example is creating an instance of the ‘environmental element’ class with a specific time/date linked to SST data as one temporal instance of Kelso Reef. Potentially this ontology could be dynamically populated in real-time, with data streaming from sensor networks and satellites.

Here, an ontology has been developed that coarsely describes the ecology of a coral reef, and the tolerance and interdependence of reef organisms like corals to physical parameters like temperature. Within the workflow, the ontology is coupled to historical datasets from the 1998 and 2002 mass bleaching events on the GBR, which are used to drive inferences between the data and ask questions for semantic correlation and analysis. Results will be used to inform an improved early warning system for coral bleaching events by inferring patterns from remotely-sensed temperature data as well as information on the existing condition of reef sites.

SEMANTIC TECHNOLOGIES

The ecosystem ontology developed and described by Myers et al. (2007) is highly complex and required validation. A sample ontology was tested against a number of simple experiments where the outcome was known. The initial tests focused predominantly on the single variable known to be the primary cause of spatially extensive or mass bleaching events – anomalous sea surface temperature. The ability to infer a bleaching alert autonomically via description logics and inference rules would verify a successful test.

The Ontology Composition

There are many different ways to engineer an ontology - the ecosystem ontology can be created in a multitude of various forms from a simple taxonomy to the complex model described here. Generally, choosing which version to create and how to fashion each object is dependent on what the ontology is designed to achieve. Here, OWL-DL and Semantic Web Rules Language (SWRL) have been used to achieve autonomy. These technologies allow the machine to make decisions based on the parameters it is given, the descriptions of the content it is storing and the datasets that are integrated (O'Connor et al., 2005). Therefore, the design decisions for this ontology aimed predominantly for a criteria of scalability, flexibility and re-usability, which are all advantages made possible with semantic technologies (Goble et al., 2004). Such criteria allowed for easier maintenance and evolution of the main ontology; to make changes, updates, etc without having to change an entire knowledge base.

Hence, the ecosystem ontology was scaled down to a basic form that included only the classes; Algae, Coral, Coral Reef, Bleached Coral Reef, Degree Heating Days and

Environmental Elements as shown in Figure 3. The rationalisation for reducing the full ecosystem ontology to this uncomplicated form for the validation process was to eliminate unforeseeable or undetectable anomalies. As this process progresses with each successful test of accuracy, complexities are again added, such as coral and algae genre.

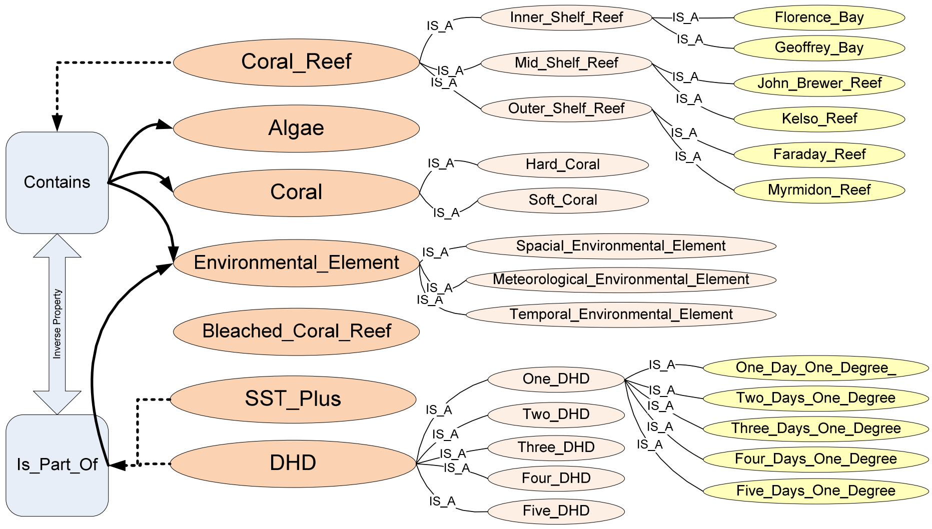

Figure 3 – A segment of the simplified Reef Ontology to be used for the validation process.

The Logic Systems

A number of differing formal logic applications are used in the validation, each orthogonal to the next. The field of Description Logics (DL) is a subset of predicate logic or first order logic (FOL) and has been used throughout history. It provides for the ability to express relations between individuals in a generalised fashion using quantification and logical axioms (Baader, 2003). For example, the property 'contains' which links instances from the domain of the 'Coral_Reef' class with instances in the range of both the 'Coral' and 'Algae' classes can be paraphrased as – 'A coral reef, among other things, contains ONLY instances belonging to Coral AND simultaneously Algae'.

Propositional logic, on the other hand, introduced by Aristotle as syllogisms, adds the ability to infer conclusions based on predefined axioms. The Semantic Web Rules Language (SWRL) uses 'horn-like' rules composed of a number of premises that contrive a logical conclusion (Horrocks et al., 2004), for example, if the first premise states 'all animals die' and the second premise states 'all corals are animals' then the logical conclusion, or consequence, can be stated as 'all corals will die'.

ASSESSING THE ONTOLOGY

In order to assess the accuracy of the ontology and inference rules a number of tests were conducted, each one slightly more complicated than the previous. The temperature indices trialled initially were the SST+ and DHDs, which are applicable to the bleaching phenomenon, among other factors (Maynard, 2004). The results showed a number of important substantiations needed to move to the next stage of this project, including the testing of not only accurate output and autonomic actions but also the scalability and flexibility of using Semantic technologies in this way. The other mitigating factors, such as water quality, salinity levels, the adaptive ability of specific coral species to water temperature, etc will be taken into account in future works as the validation process progresses.

A simple test – A bleaching alert with SST+

In the validation process, this test builds an intricacy of SWRL inference rules to analyse the SST+ anomaly, and distribute Coral Reef instances at risk of bleaching to four categories defined as properties of the 'Bleached_Coral_Reef' class in the ontology. These categories,

```
Coral_Reef(?x) ^ isSSTplusTempOf(?x, ?SSTplusAmount) ^
swrlb:greaterThanOrEqual(?SSTplusAmount, 1) ^ swrlb:lessThan(?SSTplusAmount, 2)
-> isCAT1_SSTplus(?x, true)
Coral_Reef(?x) ^ isSSTplusTempOf(?x, ?SSTplusAmount) ^
swrlb:greaterThanOrEqual(?SSTplusAmount, 2) ^ swrlb:lessThan(?SSTplusAmount, 3)
-> isCAT2_SSTplus(?x, true)
Coral_Reef(?x) ^ isSSTplusTempOf(?x, ?SSTplusAmount) ^
swrlb:greaterThanOrEqual(?SSTplusAmount, 3)
-> isCAT3_SSTplus(?x, true) ^
Bleached_Coral_Reef(?x) ^ isBleached(?x, true)
```

Figure 4 – Segment of the SWRL inference rules to categorise temporal instances of a reef's environment.

from low to high risk, will infer particular temporal instances of a reef falling within a risk area so bleach-watch alerts can be implemented.

The first set of rules automatically designates what the long-term mean temperature (LMST) is for a particular location as derived from the SST datasets for the 1997/1998 summer. Over time, this rule allows for scalability and flexibility due to local heating variables changing year by year. Hence, instead of manually inputting the data a simple modification to one inference rule can infer across the entire ontology making it more efficient.

The next set of rules takes the dynamically inferred LMST for that region (i.e. the outcome from the previous rule) and compares it against the current temperature taken at a particular time to test if there is an SST+ anomaly. If an instance of a reef, at a given time, has a current temperature reading greater than the LSMT, then determine the SST+ anomaly amount and fill the new functional data-type property; 'isSSTplusTempOf', with its value for use in the oncoming rules. The final set of rules takes the outcomes from the previous assertions and derives whether an alert is warranted by dynamically asserting instances to the four categorical levels of risk. A segment of the SWRL inference rules are shown in Figure 4.

To test the rules 9 coral instances were created, 4 Florence Bay individuals, 3 Kelso Reef individuals and 2 John Brewer Reef individuals. The outcome, shown in Figure 5, shows all instances slotted into the correct bleach watch categories. The 2 instances of interest are 'Florence_Bay_1' and 'Florence_Bay_2', with a SST+ anomaly of 3 were sent automatically to the 'Bleached_Coral_Reef' class and dynamically changing the Boolean property 'isBleached' value to true.

Figure 5 – The SST+ property assertions and inferences giving the correct results.

Building on complexity – Degree Heating Days

The Degree Heating Day (DHD) index measures the accumulation of heat stress in a given location and is an improvement on the SST anomaly as it factors in the temporal component. DHDs were defined within the ontology using DL to describe the different categories. These categories can be broken done to a finite set of rules that could be reasoned over to infer subsumption. The categories are defined as classes (refer Figure 3), 'One_DHD' to 'Five_DHDs'. The class 'One_DHD', has the only subclasses, namely 'one_day_one_degree', 'two_days_one_degree', etc, each class can be defined as being one degree above the average summer SST for multiple days. A Description Logic knowledge base consists of sets of axioms, or statements of truisms. Axioms can infer automatically one class is a subclass of another, or that an individual is an instance of an inferred class as well as its asserted class. Using quantifying axioms, incoming individual temperature data can be classified by the Reasoner to belong to an inferred class (e.i. if the temperature remains at one degree above LSMT for 2 days it can be inferred to belong to the 'Two_DHD' class). This effectively sorts the incoming data into categories that can subsequently be accessed through the inference rules. Two relatively simple ways to describe thermal stress have been used in the validation exercises presented here. There are other ways to describe thermal stress, such as the $3d_{Max}SST$ and the Heating Rate (HR) as mentioned earlier, and future works will include these and may even result in the development of improved temperature indices.

FUTURE WORK

Despite the apparent complexity of what we have described, the research, modelling, and results discussed here only represent the early stages of implementing the Semantic Reef Project in its entirety. The complexities of having a computer automatically access distributed data in a range of formats, have the machine properly interpret that data and process it to create new 'knowledge' and not just more data, are immense, but becoming more attainable.

Currently, the model is being expanded to include other temperature indices and to map dynamic data from sensors positioned on the reef, to the ontology, in order to correlate with data received daily from satellites. The inclusion of live sensor data will improve upon the spatial scale of bleaching risk forecasts currently available, and will allow the model to provide forecasts in real time.

CONCLUSION

The demand for automatic data analysis and hypothesis testing is emerging and will be made even more imperative since the infrastructure to remotely monitor reef systems using sensor networks is in the late stages of development as in the IMOS (2007) project - Great Barrier Reef Oceans Observing System (GBR-OOS). The Semantic Reef is a new approach to such data analysis and interpretation that will be extendable to other areas and issues aside from environmental monitoring and climate change.

ACKNOWLEDGEMENTS

We thank Dr Paul Marshall and Associate Professor Bruce Litow, for helpful discussions and useful feedback. We also thank the anonymous reviewer for the valuable comments and advise.

REFERENCES

Integrated Marine Observing System (IMOS): Great Barrier Reef Ocean Observing System (GBROOS) (2007). http://www.imos.org.au/nodes/great-barrier-reef-observing-system.html. Accessed June 2007.

Antoniou, G. and van Harmelen, F. (2004): *A Semantic Web Primer (Cooperative Information Systems)*, The MIT Press.

Atkinson, I. M., du Boulay, D., Chee, C., Chiu, K., Coddington, P., Gerson, A., King, T., McMullen, D., Quilici, R., Turner, P., Wendelborn, A., Wyatt, M. and Zhang, D., (2007): Developing CIMA-based cyberinfrastructure for remote access to scientific instruments and collaborative e-research. *Proceedings of the fifth Australasian symposium on ACSW frontiers - Volume 68*, Ballarat, Australia, Australian Computer Society, Inc.

Baader, F. (2003): *The Description Logic Handbook : Theory, Implementation and Applications*, Cambridge University Press.

Berkelmans, R. (2002): Time-integrated thermal bleaching thresholds of reefs and their variation on the Great Barrier Reef. *Marine Ecology Progress Series* 229 73-82.

Berkelmans, R., De'ath, G., Kininmonth, S. and Skirving, W. J. (2004): A comparison of the 1998 and 2002 coral bleaching events on the Great Barrier Reef: spatial correlation, patterns, and predictions. *Coral Reefs* 23 (1):74-83.

English, S., Wilkinson, C. and Baker, V. (1997): *Survey Manual for Tropical Marine Resources,* Townsville, Australia, ASEAN, Australia Marine Science Project, Living Coastal Resources, Australian Institute of Marine Resources.

Gleeson, M. W. and Strong, A. E. (1995): Applying MCSST to coral reef bleaching. *Advances in Space Research,* 16 (10):151-154.

Goble, C. A., De Roure, D., Shadbolt, N. R. and Fernandes, A. A. A. (2004): Enhancing services and applications with knowledge and semantics. In *The Grid 2: Blueprint for a New Computing Infrastructure,* (Eds, Foster, I. and Kesselman, C.) Morgan-Kaufmann, pp. 431-458.

Guarino, N. (1997): Understanding, building and using ontologies. *Int. J. Hum.-Comput. Stud.* 46 (2-3):293-310.

Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosof, B. and Dean, M.: SWRL: A Semantic Web Rule Language - Combining OWL and RuleML (2004). http://www.w3.org/Submission/SWRL/. Accessed July 2006.

Jennings, N. R. (2001): An agent-based approach for building complex software systems. *Communications of the ACM* 44 (4):35-41.

Kininmonth, S., Bainbridgea, S., Atkinsonc, I., Gilla, E., Barrald, L. and Vidaude, R. (2004): Sensor networking the Great Barrier Reef. *Spatial Sciences Qld* 34-38.

Marshall, P. and Schuttenberg, H. (2006): *A Reef Manager's Guide to Coral Bleaching,* Townsville, Australia, Great Barrier Reef Marine Park Authority.

Maynard, J. A. (2004): Spatial and temporal variation in bleaching susceptibility and the ability of SST variables to describe patterns in bleaching response, 1998 and 2002 event, central GBR., Thesis, School of Marine Biology, James Cook University, Townsville.

Maynard, J. A., Marshall, P. A., Johnson, J., Alves, O., Harris, G., Muller, L., Packer, G. R., Rathbone, C., Rea, A., Smith, G. P., Spillman, C., Suber, K., Tudman, P. and Turner, P. J. (2007): ReefTemp - New and improved tools to forecast and monitor coral bleaching in the Great Barrier Reef and Coral Sea. Technical report, GBRMPA, Townsville, Australia.

Myers, T. S., Atkinson, I. M. and Lavery, W. J. (2007): __The Semantic Reef: Managing Complex Knowledge to Predict Coral Bleaching on the Great Barrier Reef. In *Fifth Australasian Symposium on Grid Computing and e-Research (AusGrid 2007)*, Vol. 68 (Eds, Coddington, P. and Wendelborn, A.) ACS, Ballarat, Australia, pp. 59-67.

O'Connor, M. J., Knublauch, H., Tu, S. W., Grossof, B., Dean, M., Grosso, W. E. and Musen, M. A., (2005): Supporting Rule System Interoperability on the Semantic Web with SWRL. *Fourth International Semantic Web Conference*, Galway, Ireland, ISWC-2005.